Optimal estimators for synthetic controls

Jörg Breitung, Lennart Bolwin and Justus Töns

University of Cologne

Goethe-University Frankfurt, June 23rd, 2023

Background

- potential outcome framework developed by Neyman (1923) and refined by Rubin (1974).
- Let $Y^I_{0,t}$ be the (potential) outcome for the treated unit $i = 0$ in t
- and $Y^N_{i,t}$ is the (potential) outcome for in the absence of the intervention.
- treatment effect is defined as

$$\delta_t = Y^I_{0,t} - Y^N_{0,t}$$

- Accordingly

$$Y^I_{j,t} = \begin{cases} Y^N_{0,t} & \text{for } t \leq T_0 \\ Y^N_{0,t} + \delta_t & \text{for } t > T_0 \end{cases}$$

- goal: estimate the causal treatment effect $(\delta_{T_0}, \ldots, \delta_T)$
  counterfactual: *trajectory if there was no intervention*

Abadie/Diamond/Hainmueller (2003) approach

- Assume that there exist n additional untreated time series $Y_{1,t}, \ldots, Y_{n,t}$ for $t = 1, \ldots, T_0, \ldots, T$
- ADH assume that the counterfactual can be approximated by

$$\widehat{Y}_{0,t}^{N} \approx \sum_{i=1}^{n} w_{i,t} Y_{j,t}$$

- The weights are determined in an additional step by using vectors of k time independent pre-treatment variables $\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$
- The weights are obtained from minimizing the criterion function

$$\widehat{\boldsymbol{w}}(\boldsymbol{v}) = \operatorname*{argmin}_{\boldsymbol{w}} \left\{ \sum_{j=1}^{k} v_j \left(x_{0,j} - \sum_{i=1}^{n} w_i x_{i,j} \right)^2 \right\}$$

under the (regularisation) constraints:

$$\text{no constant} \quad 0 < w_i < 1 \quad \sum_{i=1}^{n} w_i = 1$$

Example: German reunification

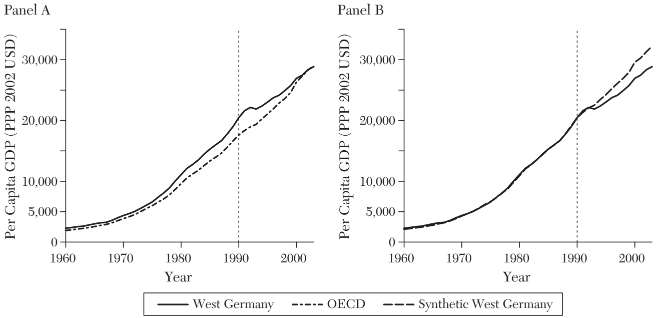

Figure 1. Synthetic Control Estimation in the German Reunification Example

Notes: Panel A compares the evolution of per capita GDP in West Germany to the evolution of per capita GDP for a simple average of OECD countries. In panel B the comparison is with a synthetic control calculated in the manner explained in subsection 3.2. See Abadie, Diamond, and Hainmueller (2015) for details.

TABLE 1
Economic Growth Predictor Means before the German Reunification

	West Germany (1)	Synthetic West Germany (2)	OECD average (3)	Austria (nearest neighbor) (4)
GDP per capita	15,808.9	15,802.2	13,669.4	14,817.0
Trade openness	56.8	56.9	59.8	74.6
Inflation rate	2.6	3.5	7.6	3.5
Industry share	34.5	34.4	33.8	35.5
Schooling	55.5	55.2	38.7	60.9
Investment rate	27.0	27.0	25.9	26.6

TABLE 2
SYNTHETIC CONTROL WEIGHTS FOR WEST GERMANY

Australia	—
Austria	0.42
Belgium	—
Denmark	—
France	—
Greece	—
Italy	—
Japan	0.16
Netherlands	0.09
New Zealand	—
Norway	—
Portugal	—
Spain	—
Switzerland	0.11
United Kingdom	—
United States	0.22

Critique

- a sound statistical framework is missing (only a rudimentary framework is provided based on a static factor model)
- the relationship between Y and $\boldsymbol{x}$ is not clear
- Consistency for $T_0 \to \infty$ and $n \to \infty$ and $n/T_0 \to 0$?
- Optimal estimator for the counterfactual?
- Leaving out the constant and the restrictions on w_i may result in biased estimates
- How to choose (v)?
- How is this approach related to the standard treatment-effect models?
- Some of these concerns are already raised in Doudchenko and Imbens (2017)

Relationship to the traditional literature on treatment-effect

- we follow to some extend Doudchenko and Imbens (2017)
- A natural estimator is the DiD-estimator that yields for $\tau > T_0$

$$\widehat{\delta}_{\tau}^{did} = \underbrace{Y_{0,\tau} - \widehat{\mu}_0^0}_{\text{a-b treated}} - \underbrace{\sum_{i=1}^{n} \left(\frac{1}{n}\right) Y_{i,\tau} - \frac{1}{n} \sum_{i=1}^{n} \left(\widehat{\mu}_i^0\right)}_{\text{a-b control}}$$

  where $\widehat{\mu}_i^0$ denote the pre-treatment sample mean of the series

- This estimator is somehow related to the ADH-approach:
  - the weights are given by $1/n$
  - no $\boldsymbol{x}_i$ is required as the weights are fixed in advance
  - The estimate are corrected for the different means of the series (in contrast to ADH)

Optimal synthetic controls: the static case

- What is the optimal estimator of δ if we assume that

$$\boldsymbol{y} = \begin{pmatrix} Y_0 \\ Y_1 \\ Y_2 \end{pmatrix} \overset{iid}{\sim} \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \text{ before } T_0$$

where $\boldsymbol{\mu} = (\mu_0, \mu_1, \mu_2)'$ and $\boldsymbol{\Sigma} \begin{pmatrix} \sigma_0^2 & \boldsymbol{\sigma}_{21}' \\ \boldsymbol{\sigma}_{21} & \boldsymbol{\Sigma}_2 \end{pmatrix}$

- Assuming that the distribution is known, the UMV estimator is given by

$$\widehat{\delta} = Y_0^I - \mathbb{E}(Y_0^N | Y_1, Y_2)$$
$$\mathbb{E}(Y_0^N | Y_1, Y_2) = \mu^* + w_1 Y_1 + w_2 Y_2$$

where $\mu^* = \mu_0 - w_1\mu_1 - w_2\mu_2$ and

$$\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \boldsymbol{\Sigma}_2^{-1} \boldsymbol{\sigma}_{21}$$

- The estimator for the CF is consistent if $n/T_0 \to 0$

Comments

- solution is similar to DiD-estimator but with different weights
- the weights depend on the correlation among the Y_i
- the DiD-estimator assumes that Y_i are equi-correlated ("time effect") and under this assumption the estimator is identical
- the optimal estimator does not impose the ADH-restrictions:
  - the constant is different from zero in general
  - the weights do not need to be positive
  - the weights do not add-up to one
- For illustration assume that

$$y \sim \mathcal{N}\left(\begin{pmatrix}1\\1\\1\end{pmatrix}, \begin{pmatrix}1 & 0.1 & 0.4\\0.1 & 1 & 0.5\\0.4 & 0.5 & 1\end{pmatrix}\right)$$

  optimal weights: $w_1 = -0.133$, $w_2 = 0.4667$.
- ADH is biased if μ_i differ and var(AHD) $=$1.16 compared to var(optimal) $= 0.827$

Estimation

- so far we assumed that the joint distribution of $Y_1, \ldots, Y_n$ is known
- OLS (plug-in) estimator:

$$Y_{0,t} = \mu^* + \sum_{i=1}^{n} w_i Y_{i,t} + u_i \quad \text{for} \quad t = 1, \ldots, T_0$$

- poor small sample properties if n/T_0 is substantial (say > 0.2)
- regularized (shrinkage) estimator minimizes

$$Q(w, \lambda_1, \lambda_2) = \sum_{t=1}^{T_0} \left(y_{0t} - \mu^* - \sum_{i=1}^{n} w_i y_{it} \right)^2 + \lambda_1 \left(\sum_{i=1}^{n} w_i^2 \right) + \lambda_2 \left(1 - \sum_{i=1}^{n} w_i \right)^2$$

  - the blue penalty term is the usual Ridge penalty
  - the red term favors weights adding up close to one
  - if λ_1 and λ_2 get large $w_i \to 1/n$
  - the minimum of Q given λ_1 and λ_2 has an explicit solution
  - similar but simpler and more natural than elastic net of DI

Factor models

- original ADH approach was motivated by a factor model
- ADH (2010) assume a perfect control with $Y_{0,t} = \sum_i w_i Y_{i,t}$
- let us assume a factor model:

$$Y_{i,t} = \mu_i + \lambda_{1i} f_{1,t} + \ldots + \lambda_{ri} f_{r,t} + \epsilon_{i,t}$$

- the optimal estimator for the synthetic control is ($r = 1$)

$$\begin{aligned}\widehat{Y}_{0t}^N &= \mathbb{E}(Y_{0t} | Y_{1,t}, \ldots, Y_{n,t}) \qquad \text{for } t = 1, 2, \ldots, T_0 \\ &= \mu_0 + \lambda_0 \, \mathbb{E}(f_t | Y_{1,t}, \ldots, Y_{n,t})\end{aligned}$$

- this suggest the following plug-in estimator:
  1. estimate the common factor(s) among $Y_{1,t}, \ldots, Y_{n,t}$ for $t = 1, \ldots, T_0$ by PCA
  2. run a regression of $Y_{0,t}$ on $(\widehat{f}_{1,t}, \ldots, \widehat{f}_{r,t})$ and a constant and denote the fitted values by $\widehat{Y}_{0,t}$
  3. run a regression $\widehat{Y}_{0,t}$ on $Y_{1,t}, \ldots, Y_{n,t}$ and a constant. The coefficients are the estimated weights
- resulting estimator is another regularized estimator where a factor structure is imposed to CovMat

Monte Carlo simulations

- Following Ferman (2021) we consider a factor model with $Y_{i,t}^{N} = \mu_i + \lambda_i' f_t + \epsilon_{i,t}$ with $r = 2$ factors
- The first and second factor loads on the treated unit and the first/second half of the donor pool.
- The loadings are all equal to one
- We simulate a total of T_0 pre-treatment and T_1 post-treatment observations
- Size of the donor pool: $n \in \{5, 10, 20, 30\}$.
- μ_i, f_{1t}, f_{2t} and ϵ_{it} are $iiN(0,1)$
  - **FACTOR**: PCA-estimator based on the factor model
  - **SC**: Synthetic-Control-Model of ADH (2010)
  - **REGOLS**: regularized OLS regression equal weight penalty
  - **NET**: regularized OLS regression with Elastic Net proposed by Doudchenko and Imbens (2017)
  - **OLS**: OLS estimator without penalty

Tabelle: Static Factor Model with $\mathbf{J = 10}$ Donors.

T_0	T_1	**FACTOR**	**SC**	**REGOLS**	**NET**	**OLS**
		RMSE	RMSE	RMSE	RMSE	RMSE
20	30	1.1613	1.2867	1.2304	1.2739	1.6198
	20	1.1555	1.2840	1.2250	1.2595	1.6167
	10	1.1312	1.2532	1.1987	1.2395	1.5811
50	30	1.1043	1.2497	1.1369	1.1467	1.2086
	20	1.0966	1.2308	1.1317	1.1407	1.2074
	10	1.0912	1.2167	1.1204	1.1281	1.1895
100	30	1.0851	1.2316	1.1063	1.1112	1.1323
	20	1.0733	1.2088	1.0917	1.0941	1.1202
	10	1.0661	1.1986	1.0895	1.0934	1.1178

Tabelle: Static Factor Model with $\mathbf{J = 30}$ Donors.

T_0	T_1	FACTOR	SC	REGOLS	NET	OLS
		RMSE	RMSE	RMSE	RMSE	RMSE
20	30	1.1013	NA	1.1899	1.2523	NA
	20	1.1043	NA	1.1825	1.2442	NA
	10	1.0890	NA	1.1647	1.2337	NA
50	30	1.0622	1.1477	1.1039	1.1297	1.6851
	20	1.0508	1.1372	1.0936	1.1190	1.6716
	10	1.0445	1.1342	1.0880	1.1061	1.6681
100	30	1.0420	1.1178	1.0662	1.0793	1.2394
	20	1.0435	1.1237	1.0669	1.0788	1.2351
	10	1.0159	1.0890	1.0410	1.0465	1.2026


MSFE-Mean
1.4
1.2
10
20
30
Number of Donors
type
POST_SC_RMSFE
POST_OLS_RMSFE
POST_REGOLS_RMSFE
POST_NET_RMSFE
POST_FACTOR_RMSFE

Dynamic models

- Assume that the $(k+1) \times 1$ vector of time series $\boldsymbol{y}_t = (Y_{0,t}, \ldots, Y_{n,t})'$ can be represented as

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{\alpha} + \boldsymbol{A}_1 \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{A}_p \boldsymbol{y}_{t-p} + \boldsymbol{u}_t \\ \boldsymbol{y}_t &= \boldsymbol{\mu} + \boldsymbol{A}(L)(\boldsymbol{y}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{u}_t \end{aligned}$$

  where $\boldsymbol{\mu} = (\mu_0, \mu_1, \ldots, \mu_n)'$, $\mathbb{E}(\boldsymbol{u}_t \boldsymbol{u}_t') = \boldsymbol{\Sigma}$ is a positive definite covariance matrix.
- The counterfactual is given by

$$\begin{aligned} \widehat{Y}_{0,t}^N &= \mathbb{E}(Y_{0,t}^N | \mathcal{I}_t) \\ &= \mu_0 + \sum_{i=1}^{n} w_i (Y_{i,t} - \mu_i) + \beta(L)'(\widetilde{y}_{t-1} - \mu) \end{aligned}$$

  where $(\widetilde{y}_{t-1} - \mu) = (\widehat{Y}_{0,t}^N - \mu_0, Y_{1,t} - \mu_1, \ldots, Y_{n,t} - \mu_n)'$
- lags of the series enter the estimated counterfactual
- $\widehat{Y}_{0,t}^N$ needs to be estimated recursively

A univariate representation

- using $\widehat{Y}^N_{0,t} = \boldsymbol{w}'\boldsymbol{y}_t$ we obtain the univariate representation:

$$\boldsymbol{y}^N_{0,t} = \alpha_0 + \sum_{j=0}^{p} \alpha_j \left(\boldsymbol{w}'\boldsymbol{y}_{t-j}\right) + e_t \tag{1}$$

- estimate the parameters by sequential OLS:
  - select some initial guess for $\boldsymbol{w}$ and estimate α_j in (1) by OLS
  - fix α_j and estimate $\boldsymbol{w}$ by running the regression

$$\boldsymbol{y}^N_{0,t} = \alpha_0 + \boldsymbol{w}' \underbrace{\left(\sum_{j=0}^{p} \widehat{\alpha}_j \boldsymbol{y}_{t-j} \right)}_{\boldsymbol{z}_t} + e_t$$

    which yields the updated weights
  - repeat the estimation until convergence
- Alternatively some regularized estimator for the full dynamic model may be used.

Conclusions

- Synthetic control methods may be useful for assessing the causal effect of economic interventions
- if the donor series are correlated with the treated series, then the weights are different from $(1/n)$ and need~~s~~ to be estimated
- we derive the optimal (OLS) estimator for the counterfactual in a simple static setup
- we argue that the weights do not need to satisfy the adding-up constraint, not do they need to be positive
- it is important to include a constant term when estimating the weights
- if n/T_0 is substantial, some regularization is required
- we propose some penalization that shrinks the estimates towards $1/n$
- for dynamic models, the optimal counterfactual involves the lags of the donor series
- we propose a simple univariate autoregressive representation for the counterfactual